REMOTE IMAGE CAPTURE
WITH CENTRALIZED PROCESSING AND STORAGE

CROSS-REFERENCE TO RELATED APPLICATIONS

This application is a continuation in part of application Serial No. 08/917,761 filed August 27, 1997, now U.S. Patent 5,910,988.

FIELD OF THE INVENTION

This invention relates generally to the automated processing of documents and electronic data from different applications including sale, business, banking and general consumer transactions. More particularly, it pertains to an automated system to retrieve transaction data at remote locations, to encrypt the data, to transmit the encrypted data to a central location, to transform the data to a usable form, to generate informative reports from the data and to transmit the informative reports to the remote locations.

BACKGROUND

This invention involves the processing of documents and electronic data which are generated, for example, from sale, business and banking transactions including credit card transactions, smart card transactions, automated teller machine (ATM) transactions, consumer purchases, business forms, W2 forms, birth certificates, deeds and insurance documents.

The enormous number of paper and electronic records generated from documents and electronic data from sale, business and banking transactions contain valuable information. First, these paper and electronic records contain information which can be used to verify the accuracy of the records maintained by consumers, merchants and bankers. For example, customers use paper receipts of sale and banking transactions to verify the information on the periodic statements which they receive from their bank or credit card institution. Merchants use paper receipts to record sale transactions for management of customer complaints. Taxpayers use paper receipts to record tax deductible contributions for use in their tax return preparation. Employees use paper receipts to record business expenses for preparation of business expense forms.

Paper and electronic records also contain information which can be used for market analysis. For example, manufacturers and retailers can determine consumer preferences in different regions as well as trends in consumer preferences from the information contained in paper and electronic records.

However, the maintenance and processing of paper and electronic records presents difficult challenges. First, paper receipts and documents could easily be lost, misplaced, stolen, damaged or destroyed. Further, the information contained in these paper and electronic records cannot be easily processed because it is scattered among individual records. For example, the market trend information contained in a group of sales records retained by merchants cannot easily be determined since this information is scattered among the individual records. Likewise, the tax information contained in a group of paper receipts of sales transactions retained by consumers cannot easily be processed.

Previous approaches have been proposed to meet the challenges associated with the maintenance and processing of paper and electronic records. For example, data archive service companies store the information from paper receipts and documents acquired from their customers on microfilm or compact disc read only memory (CD-ROM) at a central facility. Customers typically deliver the paper receipts and documents to the central facility. For sensitive documents which cannot leave the customer site, some data archive service companies perform data acquisition and transfer to magnetic tapes at the customer site and deliver the tapes to the central facility.

The approach offered by these data archive service companies have disadvantages. First, the approach is costly and has poor performance because it requires an expensive, time consuming physical transportation of paper receipts or magnetic tapes from the customer site to the central facility. Further, the approach is not reliable as information can be lost or damaged during physical

transportation. The approach also has limited capability as it does not process electronic records along with the paper receipts within a single system.

Other approaches have focused on the elimination of paper receipts and documents. U.S. Patent No. 5,590,038 discloses a universal electronic transaction card (UET card) or smart card which stores transaction information on a memory embedded on the card as a substitute for a paper receipt. Similarly, U.S. Patent No. 5,479,510 discloses a method of electronically transmitting and storing purchaser information at the time of purchase which is read at a later time to ensure that the purchased goods or services are delivered to the correct person.

While these approaches avoid the problems associated with paper receipts, they have other disadvantages. First, these approaches do not offer independent verification of the accuracy of the records maintained by consumers, merchants and bankers with a third party recipient of the transaction data. For example, if a UET card is lost, stolen, damaged or deliberately altered by an unscrupulous holder after recording sale or banking transactions, these approaches would not be able to verify the remaining records which are maintained by the other parties to the transactions.

Next, these approaches do not have the ability to process both paper and electronic records of transactions within a single, comprehensive system. Accordingly, they do not address the task of processing the enormous number of paper receipts which have been generated from sales and banking transactions. The absence of the ability to process both paper and electronic records of these approaches is a significant limitation as paper receipts and documents will continue to be generated for the foreseeable future because of concerns over the reliability and security of electronic transactions and the familiarity of consumers and merchants with paper receipts.

These approaches also have a security deficiency as they do not offer signature verification which is typically used

on credit card purchases to avoid theft and fraud. For example, a thief could misappropriate money from a UET card holder after obtaining by force, manipulation or theft the user's personal identification number (PIN). Similarly, it is not uncommon for criminals to acquire credit cards in victims' names and make unlawful charges after obtaining the victim's social security number. This becomes a greater concern as that type of personal information becomes available, e.g., on the internet. Also, the signature verification performed manually by merchants for credit card purchases frequently misses forged signatures.

Even if smart cards or UET cards had the ability to store signature and other biometric data within the card for verification, the system would still have disadvantages. First, the stored biometric data on the card could be altered by a card thief to defeat the security measure. Similarly, the biometric data could be corrupted if the card is damaged. Finally, the security measure would be costly at it would require an expensive biometric comparison feature either on each card or on equipment at each merchant site.

Additional biometric verification systems including signature verification systems have been proposed to address the security problem. For example, U.S. Patent 5,657,393 discloses a method and apparatus for verification of hand-written signatures involving the extraction and comparison of signature characteristics including the length and angle of select component lines. In addition, U.S. Patent 5,602,933 discloses a method and apparatus for the verification of remotely acquired data with corresponding data stored at a central facility.

However, none of these verification systems offer general support for transaction initiation, remote paper and electronic data acquisition, data encryption, data communication , data archival, data retrieval, data mining, manipulation and analytic services. Accordingly, there is a need for a single system which offers comprehensive support for the tasks involved in the automated processing of

5

documents, biometric and electronic data from sale, business, banking and general consumer transactions. Further, there is a need for a single comprehensive system having the reliability, performance, fault tolerance, capacity, cost and security to satisfy the requirements of the retail, business, banking and general consumer industries.

SUMMARY OF THE INVENTION

The invention provides an automated, reliable, high performance, fault tolerant, and low cost system with maximal security and availability to process electronic and paper transactions, and has been named the DataTreasury™ System.

It is an object of the present invention to provide a system for central management, storage and verification of remotely captured electronic and paper transactions from credit cards, smart cards, debit cards, documents and receipts involving sales, business, banking and general purpose consumer applications comprising:

at least one remote data access subsystem for capturing and sending electronic and paper transaction data;

at least one data collecting subsystem for collecting and sending the electronic and paper transaction data comprising a first data management subsystem for managing the collecting and sending of the transaction data;

at least one central data processing subsystem for processing, sending and storing the electronic and paper transaction data comprising a second data management subsystem for managing the processing, sending and storing of the transaction data; and

at least one communication network for the transmission of the transaction data within and between said at least one data access subsystem and said at least one data processing subsystem.

The DataTreasury™ System processes paper and/or electronic receipts such as credit card receipts, Automated Teller Machine (ATM) receipts, business expense receipts and sales receipts and automatically generates reports such as

credit card statements, bank statements, tax reports for tax return preparation, market analyses, and the like.

It is a further object of the DataTreasury™ System to retrieve both paper and electronic transactions at remote locations.

It is a further object of the DataTreasury™ System to employ a scanner and a data entry terminal at a customer site to retrieve data from paper transactions and to enable additions or modifications to the scanned information respectively.

It is a further object of the DataTreasury™ System to provide an input device for retrieving transaction data from the memory of smart cards for independent verification of the records maintained by consumers, merchants and bankers to prevent the loss of data from the loss, theft, damage or deliberate alteration of the smart card.

It is a further object of the DataTreasury™ System to retrieve and process transaction data from DataTreasury™ System anonymous smart cards which are identified by an account number and password. Since DataTreasury™ System anonymous smart card transactions can be identified without the customer's name, a customer can add money to the DataTreasury™ System anonymous smart card and make expenditures with the card with the same degree of privacy as cash acquisitions and expenditures.

It is a further object of the DataTreasury™ System to retrieve customer billing data from employee time documents and to generate customer billing statements from the billing data.

It is a further object of the DataTreasury™ System to initiate electronic transactions including transactions on the internet and to provide identification verification by capturing and comparing signature and biometric data.

It is a further object of the DataTreasury™ System of the invention to process electronic and paper transactions with a tiered architecture comprised of DataTreasury™ System

Access Terminals (DATs), DataTreasury™ System Access Collectors (DACs) and DataTreasury™ System Processing Concentrators (DPCs).

BRIEF DESCRIPTION OF THE DRAWINGS

These and other objects and features of the invention will be more clearly understood from the following detailed description along with the accompanying drawing figures, wherein:

FIG. 1 is a block diagram showing the three major operational elements of the invention: the DataTreasury™ System Access Terminal (DAT), the DataTreasury™ System Access Collector (DAC) and the DataTreasury™ System Processing Concentrator (DPC);

FIG. 2 is a block diagram of the DAT architecture;

FIG. 3a is a flow chart describing image capture by a DAT;

FIG. 3b displays a sample paper receipt which is processed by the DAT;

FIG. 4 is a block diagram of the DAC architecture;

FIG. 5 is a flow chart describing the polling of the DATs by a DAC;

FIG. 6 is a block diagram of the DPC architecture;

FIG. 7 is a flow chart describing the polling of the DACs by the DPC;

FIG. 8 is a flow chart describing the data processing performed by the DPC; and

FIG. 9 is a flow chart describing the data retrieval performed by the DPC; and

FIG. 10 is a flow chart describing the use of the DataTreasury™ system to process personal checks.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

FIG. 1 shows the architecture of the DataTreasury™ System 100. The DataTreasury™ System 100 has three operational elements: the DataTreasury™ System Access Terminal (DAT) 200 (the remote data access subsystem), the

DataTreasury™ System Access Collector (DAC) 400 (the intermediate data collecting subsystem), and the DataTreasury™ System Processing Concentrator (DPC) 600 (the central data processing subsystem).

The DataTreasury™ System 100 architecture consists of three tiers. At the bottom tier, the DATs 200 retrieve data from the customer sites. At the next tier, the DACs 400 poll the DATs 200 to receive data which accumulates in the DATs 200. At the top tier, the DPCs 600 poll the DACs 400 to receive data which accumulates in the DACs 400. The DPCs 600 store the customer's data in a central location, generate informative reports from the data and transmit the informative reports to the customers at remote locations.

In the preferred embodiment, the DataTreasury™ System 100 complies with the Price Waterhouse SAS70 industry standard. Specifically, the DataTreasury™ System 100 meets the software development standard, the system deployment standard and the reliability standard specified by Price Waterhouse SAS70. By adhering to the Price Waterhouse SAS70 standard, the DataTreasury™ System 100 provides the security, availability and reliability required by mission critical financial applications of banks and stock brokerage companies.

As is known to persons of ordinary skill in the art, the DataTreasury™ System 100 could also use other software development standard, other system deployment standards and other reliability standards as long as adherence to these alternative standards provides the security, availability and reliability required by mission critical financial applications.

FIG. 2 shows a block diagram of the DAT 200 architecture. DATs 200 are located at customer sites. The DataTreasury™ System 100 customers include merchants, consumers and bankers. The DATs 200 act as the customer contact point to the suite of services provided by the DataTreasury™ System 100. In the preferred embodiment, the DAT 200 is custom designed around a general purpose thin

9

client Network Computer (NC) which runs SUN Microsystem's JAVA/OS operating system. The custom designed DAT 200 comprises a DAT scanner 202, a DAT modem 204, DAT digital storage 206, a DAT controller 210 (workstation), a DAT card interface 212, an optional DAT printer 208 and a signature pad 214.

As is known to persons of ordinary skill in the art, the DAT 200 could also be custom designed around a general purpose network computer running other operating systems as long as the chosen operating system provides support for multiprocessing, memory management and dynamic linking required by the DataTreasury™ System 100.

The DAT scanner 202 scans a paper receipt and generates a digital bitmap image representation called a Bitmap Image (BI) of the receipt. In the preferred embodiment, the DAT scanner 202 has the ability to support a full range of image resolution values which are commonly measured in Dots Per Inch (DPI). Next, the DAT scanner 202 has the ability to perform full duplex imaging. With full duplex imaging, a scanner simultaneous captures both the front and back of a paper document. The DAT scanner 202 can also support gray scale and full color imaging at any bit per pixel depth value. The DAT scanner 202 also supports the capture of hand-written signatures for identity verification.

In addition to scanning images and text, the DAT scanner 202 also scans DataGlyph™ elements, available from Xerox Corporation. As is known to persons of ordinary skill in the art, the Xerox DataGlyph™ Technology represents digital information with machine readable data which is encoded into many, tiny, individual glyph elements. Each glyph element consists of a 45 degree diagonal line which could be as short as 1/100th of an inch depending on the resolution of the scanning and printing devices. Each glyph element represents a binary 0 or 1 depending on whether it slopes downward to the left or the right respectively. Accordingly, DataGlyph™ elements can represent character strings as ASCII or EBCIDIC binary representations. Further, encryption methods, as

known to persons of ordinary skill in the art encrypt the data represented by the DataGlyph™ Technology.

The use of glyph technology in the DataTreasury™ System 100 improves the accuracy, cost and performance of the system. Xerox DataGlyph™ Technology includes error correction codes which can be referenced to correct scanning errors or to correct damage to the document caused by ink spills or ordinary wear. DataGlyph™ Technology also leads to decreased system cost since the system will require less manual intervention for data entry and correction because of the improved accuracy associated with DataGlyph™ elements. Since DataGlyph™ elements represent a large amount of information in a small amount of space, the DAT scanner 100 will require a small amount of time to input a large amount of information.

The DAT card interface 212 and the DAT signature pad 214 along with the internet and telephone access through the DAT modem 204 enable the DataTreasury™ System 100 customer to initiate secure sale and banking transactions via the internet or telephone with the DAT 200 using a variety of cards including debit cards, smart cards and credit cards. After selecting a purchase or a banking transaction through a standard internet interface, the DataTreasury™ System 100 customer inserts or swipes the debit card, smart card or credit card into the DAT card interface 212.

The DAT card interface 212 retrieves the identification information from the card for subsequent transmission to the destination of the internet transaction. Further, the DAT scanner 202 could capture a hand written signature from a document or the DAT signature pad 214 could capture an electronic signature written on it with a special pen. Similarly, these security features allow a credit card recipient to activate the card with a DAT 200 located at a merchant site. The security features would detect unauthorized use of debit cards, credit cards and smart cards resulting from their unlawful interception. Accordingly, the DataTreasury™ System's 100 security features offer a more

secure alternative for internet and telephone transactions than the typical methods which only require transmission of a card account number and expiration date.

As is known to persons of ordinary skill in the art, the DATs 200 could also include additional devices for capturing other biometric data for additional security. These devices include facial scans, fingerprints, voice prints, iris scans, retina scans and hand geometry.

In addition to initiating sale and banking transactions, the DAT card interface 212 also reads sale and banking transactions initiated elsewhere from the memory of smart cards to enable subsequent storage and processing by the DataTreasury™ System. If a smart card is lost, stolen, damaged or deliberately altered by an unscrupulous holder after the DAT card interface 212 reads its transaction data, the DataTreasury™ System 100 can reproduce the transaction data for the customer. Accordingly, the DAT card interface 212 provides support for independent verification of the records maintained by consumers, merchants and bankers to prevent the loss of data from the loss, theft, damage or deliberate alteration of the smart card.

The DAT card interface 212 also supports the initiation and retrieval of sale and banking transactions with the DataTreasury™ System anonymous smart cards. In contrast to standard debit cards and credit cards, the DataTreasury™ System anonymous smart card does not identify the card's holder by name. Instead, the DataTreasury™ System anonymous smart card requires only an account number and a password. Since DataTreasury™ System anonymous smart card transactions can be identified without the customer's name, a DataTreasury™ System 100 customer can purchase a DataTreasury™ System anonymous smart card, add money to the card, make expenditures with the card and monitor the card's account with the same degree of privacy as cash acquisition, expenditure and management.

The DAT scanner 202, the internet access, the signature pad 214 and other biometric data capture devices also support

the remote capture of survey information and purchase orders. For example, the DAT scanner 202 captures surveys appearing on the back of checks at restaurants and bars. Similarly, the DAT scanner 202 could capture purchase orders from residences, enabling customers to make immediate purchases from their home of goods promoted through the mail. Accordingly, home marketing merchant could transmit sales in a more cost efficient and reliable manner by using the DAT scanner 202 instead of providing envelopes with prepaid postage to residences.

The DAT scanner 202 also captures receipts which are subsequently needed for tax return preparation or tax audits. Similarly, the DAT scanner 202 captures sales receipts from merchants, providing an off-site secure, reliable repository to guard against loss resulting from flooding, fire or other circumstances. This feature could also allow a merchant to automatically perform inventory in a reliable and cost-effective manner.

The DAT controller 210 performs processing tasks and Input/Output (I/O) tasks which are typically performed by a processor. The DAT controller 210 compresses, encrypts and tags the BI to form a Tagged Encrypted Compressed Bitmap Image (TECBI). The DAT controller 210 also manages the Input/Output (I/O). Specifically, the DAT controller 210 manages devices like the DAT scanner 202, the DAT digital storage 206, the optional DAT printer 208 and the DAT modem 204.

The DAT digital storage 208 holds data such as the TECBI. The DAT modem 204 transmits data from the DAT 200 to the appropriate DAC 400 as instructed by the DAT controller 210. Specifically, the DAT modem 204 transmits the TECBIs from the DAT digital storage 208 to the appropriate DAC 400. In the preferred embodiment, the DAT modem 204 is a high speed modem with dial-up connectivity. The DAT digital storage 208 is sufficiently large to store the input data before transmission to a DAC 400. The DAT digital storage 208 can be Random Access Memory (RAM) or a hard drive.

FIG. 3a is a flow chart 300 describing the operation of the DAT in detail. In step 310, the DAT scanner 202 scans paper receipts into the DAT 200 provided by an operator. In step 312, the DAT controller 210 determines whether the operation executed successfully. If the scanning is successful, the DAT scanner 202 produces a Bitmap Image (BI). If the scanning is unsuccessful, the DAT controller 210 notifies the operator of the trouble and prompts the operator for repair in step 370.

If a BI is created, the DAT controller 210 executes a conventional image compression algorithm like the Tagged Image File Format (TIFF) program to compress the BI in step 314. In step 316, the DAT controller 210 determines whether the compression executed successfully. If the compression is successful, it produces a Compressed Bitmap Image (CBI). If the compression is unsuccessful, the DAT controller 210 notifies the operator of the trouble and prompts the operator for repair in step 370.

If a CBI is created, the DAT controller 210 executes an encryption algorithm which is well known to an artisan of ordinary skill in the field to encrypt the CBI in step 318. Encryption protects against unauthorized access during the subsequent transmission of the data which will be discussed below. In step 320, the DAT controller 210 determines whether the encryption operation executed successfully. If the encryption is successful, it produces an Encrypted Compressed Bitmap Image (ECBI). If the encryption is unsuccessful, the DAT controller 210 notifies the operator of the trouble and prompts the operator for repair in step 370.

If an ECBI is created, the DAT controller 210 tags the ECBI with a time stamp which includes the scanning time, an identification number to identify the merchant originating the scan and any additional useful information in step 322. In step 324, the DAT controller 210 determines whether the tagging operation executed successfully. If the tagging is successful, it produces a Tagged Encrypted Compressed Bitmap Image (TECBI). If the tagging is unsuccessful, the DAT

controller 210 notifies the operator of the trouble and prompts the operator for repair in step 370.

If a TECBI is created, the DAT controller 210 stores the TECBI in the DAT digital storage 208 in step 326. In step 328, the DAT controller 210 determines whether the storing operation executed successfully. If the storing operation is successful, the DAT digital storage 208 will contain the TECBI. If the storing operation is unsuccessful, the DAT controller 210 notifies the operator of the trouble and prompts the operator for repair in step 370.

If the TECBI is properly stored in the DAT digital storage 208, the DAT controller 210 determines whether all paper receipts have been scanned in step 330. If all paper receipts have not been scanned, control returns to step 310 where the next paper receipt will be processed as discussed above. If all paper receipts have been scanned, the DAT controller 210 asks the operator to verify the number of scanned receipts in step 334. If the number of scanned receipts as determined by the DAT controller 210 does not equal the number of scanned receipts as determined by the operator, the DAT controller 210 asks whether the operator desires to rescan all of the receipts in step 338.

If the operator chooses to rescan all of the receipts in step 338, the DAT controller 210 will delete all of the TECBIs associated with the batch from the DAT digital storage 208 in step 342. After the operator prepares the batch of receipts for rescan in step 346, control returns to step 310 where the first receipt in the batch will be processed as discussed above.

If the operator chooses not to rescan all of the receipts from the batch in step 338, control returns to step 334 where the DAT controller 210 asks the operator to verify the number of scanned receipts as discussed above.

If the number of scanned receipts as determined by the DAT controller 210 equals the number of scanned receipts as determined by the operator, the DAT controller 210 prints a batch ticket on the DAT printer 206 in step 350. The

PEDC-93965.3

operator will attach this batch ticket to the batch of receipts which have been scanned. This batch ticket shall contain relevant session information such as scan time, number of receipts and an identification number for the data operator. If processing difficulties occur for a batch of receipts after the image capture of flowchart 300, the batch ticket will enable them to be quickly located for rescanning with the DAT 200.

In step 354, the DAT controller 210 determines whether the scan session has completed. If the scan session has not completed, control returns to step 310 where the first receipt in the next batch of the scan session will be processed as discussed above. If the scan session has completed, the DAT controller 210 selectively prints a session report on the DAT printer 206 in step 358. The DAT controller 210 writes statistical information for the session to the DAT digital storage 208 in step 362. In step 366, the DAT controller 210 terminates the session.

FIG. 3b displays a sample paper receipt which is processed by the DAT 200 as described by the flowchart in FIG. 3a. The sample paper receipt involves a credit card transaction which has four participants:

A. The *ISSUER*: is an entity such as a bank or corporate financial institution such as GE Capital, GM or AT&T which provides the credit behind the credit card and issues the card to the consumer.

B. The *PROCESSOR*: executes the processing of an inbound credit card transaction by performing basic transaction validation that includes checking with the ISSUER database to ensure that the credit card has sufficient credit to allow approval of the transaction.

C. The *ACQUIRER*: specializes in the marketing, installation and support of Point Of Sale (POS) credit card terminals. The acquirer, like the DAC 400 in the DataTreasury™ System 100 acts as an electronic collection point for the initial credit card transaction as the card is

16

inserted into the POS terminal. After acquisition, the acquirer passes the transaction to the PROCESSOR.

D. <u>The *MERCHANT*</u>: inserts a credit card into a POS terminal and enters the amount of the transaction to initiate the credit card transaction.

In the preferred embodiment, the DAT 200 reads the following information from the sample paper receipt shown in FIG. 3b and stores the information in the format described below.

CUSTOMER_ID 370 : This field is a 7 position HEX numeric value. This field uniquely identifies the customer using the terminal. In this sample, this field would identify the credit card merchant.

TERMINAL_ID 372: This field is a 6 position decimal numeric value. This field uniquely identifies the credit card terminal which is used to print the credit card receipt.

TRANSACTION_DATE 374: This field contains the date and time of the credit card transaction.

TRANSACTION_LINE_ITEM 376: This field is a variable length character string. The first three positions represent a right justified numeric field with leading zeros indicating the full length of this field. This field contains all data pertaining to the purchased item including the item's price. The DAT 200 will store a TRANSACTION_LINE_ITEM field for each transaction line item on the receipt. This field is optional since not all credit card transactions will have line items.

TRANSACTION_SUBTOTAL 378: This field is a double precision floating point number. This field indicates the subtotal of the TRANSACTION_LINE_ITEMs.

TRANSACTION_SALES_TAX 380: This field is a double precision floating point number. This field contains the sales tax of the TRANSACTION_SUBTOTAL.

TRANSACTION_AMOUNT 382: This field is a double precision floating point number. This field is the sum of the TRANSACTION_SUBTOTAL and TRANSACTION_SALES_TAX.

CREDIT_CARD_ACCT_NUM 384: This field is a 12 position decimal value. This field identifies the credit card which was used to execute this transaction.

CREDIT_CARD_EXP_DATE 386: This field identifies the expiration date of the credit card.

TRANSACTION_APPROVAL_CODE 388: This field is a 6 position numeric value. This field indicates the approval code that was given for the particular transaction.

The DAT 200 also stores additional items which are not pictured in FIG. 3b as described below:

ISSUER_ID: This field is a 7 position decimal numeric value. This field identifies the credit card issuer.

ACQUIRER_ID: This field is a 7 position decimal numeric value. This field identifies the acquirer.

PROCESSOR_ID: This field is a 7 position decimal numeric value. This field identifies the processor.

TRANSACTION_LINE_ITEM_CNT: This field is a 3 position decimal numeric value. This field identifies the number of transaction line items on the receipt. A value of ZERO indicates the absence of any transaction line items on the receipt.

TRANSACTION_GRATUITY: This field is a double precision floating number. This field is optional because it will only appear on restaurant or bar receipts.

FINAL_TRANSACTION_AMOUNT: This field is a double precision floating number. This field is optional because it will only appear on restaurant and bar receipts. The field is the sum of the TRANSACTION_AMOUNT and TRANSACTION_GRATUITY.

The tag prepended to the ECBI in step 322 of the flowchart of FIG. 3a identifies the time and place of the document's origination. Specifically, the tag consists of the following fields:

DAT_TERMINAL_ID: This field is a 7 position hexadecimal numeric value. This field uniquely identifies the DAT 200 which is used by the customer.

DAT_SESSION_DATE: This field identifies the date and time of the DAT 200 session which generated the image of the document.

DAT_USER_ID: This field is a 4 position decimal numeric value. This field identifies the individual within the CUSTOMER's organization who initiated the DAT 200 session.

DATA_GLYPH_RESULT: This field is a variable length character string. The first four positions hold a right justified numeric position with leading zero which indicate the length of the field. The fifth position indicates the DataGlyph™ element status. A value of 0 indicates that the data glyph was NOT PRESENT on the receipt. A value of 1 indicates that the data glyph WAS PRESENT and contained no errors. A value of 2 indicates that the data glyph WAS PRESENT and had nominal errors. If the fifth position of this field has a value of 2, the remaining portion of the string identifies the erroneous field numbers. As subsequently described, the DPC 600 will reference this portion of the field to capture the erroneous data from the receipt with alternate methods. A value of 3 indicates that the data glyph WAS PRESENT WITH SEVERE ERRORS. In other words, a value of 3 indicates the DataGlyph™ element was badly damaged and unreadable.

The receipt shown in FIG. 3b can also contain a signature which can be captured by the DAT scanner 202. A data glyph could identify the location of the signature on the receipt.

As is known to persons of ordinary skill in the art, the DataTreasury™ System 100 can also process receipts with alternate formats as long as the receipt contains the appropriate identification information such as the transaction amount, the customer, the DAT 200, the transaction date, the transaction tax, the credit card number, the credit card expiration date, etc.

The DataTreasury™ System 100 partitions the paper receipt into image snippets as illustrated by the sample on FIG. 3b. Partitioning facilitates an improvement in the

process to correct errors from the scanning operation. If an error occurred during scanning, the DataTreasury™ System 100 corrects the error using manual entry. With partitioning, the DataTreasury™ System 100 focuses the correction effort on only the image snippet having the error instead of correcting the entire document. The subsequently discussed schema of the DataTreasury™ System 100 database describes the implementation of the partitioning concept in detail.

The DACs 400 form the backbone of the tiered architecture shown in FIG. 1 and FIG. 4. As shown in FIG. 1, each DAC 400 supports a region containing a group of DATs 200. Each DAC 400 polls the DATs 200 in its region and receives TECBIs which have accumulated in the DATs 200. The DACs 400 are located at key central sites of maximum merchant density.

In the preferred embodiment, the DAC server 402 comprises stand-alone Digital Equipment Corporation (DEC) SMP Alpha 4100 2/566 servers which are connected on a common network running Windows NT. The DEC Alpha servers manage the collection and intermediate storage of images and data which are received from the DATs 200.

As is known to persons of ordinary skill in the art, the DataTreasury™ System 100 could use any one of a number of different servers that are available from other computer vendors as long as the server meets the capacity, performance and reliability requirements of the system.

In the preferred embodiment, the DAC server 402 also comprises EMC 3300 SYMMETRIX CUBE Disk Storage Systems, which store the images and data collected and managed by the DEC Alpha servers. The DAC 400 architecture also uses a SYMMETRIX Remote Data Facility (SRDF), available from EMC, to enable multiple, physically separate data centers housing EMC Storage Systems to maintain redundant backups of each other across a Wide Area Network (WAN). Since SRDF performs the backup operations in the background, it does not affect the operational performance of the DataTreasury™ System 100. The DAC server 402 also has secondary memory 410. In the

preferred embodiment, the secondary memory 410 is a small scale DLT jukebox.

The DAC Alpha servers of the DAC server 402 insert images and data received from the DATs 200 into a database which is stored on the disk storage systems using a data manipulation language as is well known to persons of ordinary skill in the art. In the preferred embodiment, the database is a relational database available from Oracle.

As is well known to persons of ordinary skill in the art, the DataTreasury™ System 100 could use any one of a number of different database models which are available from other vendors including the entity relationship model as long as the selected database meets the storage and access efficiency requirements of the system. See, e.g., Chapter 2 of Database System Concepts by Korth and Silberschatz.

The DAC 400 architecture uses a WEB based paradigm using an enhanced Domain Name Services (DNS), the Microsoft Component Object Model (DCOM), and Windows NT Application Program Interfaces (APIs) to facilitate communication and load balancing among the servers comprising the DAC server 402. As is known to persons of ordinary skill in the art, DNS, which is also known as Bind, statically translates name requests to Internet Protocol 4 (IP4) addresses. In the DAC 400 architecture, an enhanced DNS dynamically assigns IP4 addresses to balance the load among the servers comprising the DAC server 402.

In the preferred embodiment, the enhanced DNS is designed and implemented using objects from Microsoft DCOM. Using the DCOM objects, the enhanced DNS acquires real-time server load performance statistics on each server comprising the DAC server 402 from the Windows NT API at set intervals. Based on these load performance statistics, the enhanced DNS adjusts the mapping of name requests to IP4 addresses to direct data toward the servers which are more lightly loaded.

A large bank of modems 404 polls the DATs 200 at the customer sites within the DAC's 400 region. In the preferred embodiment, the bank of modems 404, available as CISCO

AS5200, is an aggregate 48 modem device with Local Area Network (LAN) 406 connectivity which permits the DAC servers 402 to dial the DATs 200 without requiring 48 separate modems and serial connections.

The DAC servers 402 and the bank of modems 404 are connected on a LAN 406. In the preferred embodiment, the LAN uses a switched 100BaseT/10BaseT communication hardware layer protocol. As is known to persons of ordinary skill in the art, the 100BaseT/10BaseT protocol is based on the Ethernet model. Further, the numbers 100 and 10 refer to the communication link speed in megabits per second. In the preferred embodiment, the CISCO Catalyst 2900 Network Switch supports the LAN 406 connectivity between the devices connected to the LAN 406 including the DAC servers 402 and the bank of modems 404.

As is known to persons of ordinary skill in the art, alternate LAN architectures could be used to facilitate communication among the devices of the LAN 406. For example, the LAN 406 could use a hub architecture with a round robin allocation algorithm, a time division multiplexing algorithm or a statistical multiplexing algorithm.

A Wide Area Network (WAN) router 408 connects the LAN 406 to the WAN to facilitate communication between the DACs 400 and the DPCs 600. In the preferred embodiment, the WAN router 408 is a CISCO 4700 WAN Router. The WAN router 408 uses frame relay connectivity to connect the DAC LAN 406 to the WAN. As is known to persons of ordinary skill in the art, alternate devices, such as the NORTEL Magellen Passport "50" Telecommunication Switch, could be used to facilitate communication between the DACs 400 and the DPCs 600 as long as the selected router meets the performance and quality communication requirements of the system.

As is known to persons of ordinary skill in the art, frame relay is an interface protocol for statistically multiplexed packet-switched data communications in which variable-sized packets (frames) are used that completely enclose the user packets which they transport. In contrast

to dedicated point to point links that guarantee a specific data rate, frame relay communication provides bandwidth on-demand with a guaranteed minimum data rate. Frame relay communication also allows occasional short high data rate bursts according to network availability.

Each frame encloses one user packet and adds addressing and verification information. Frame relay data communication typically has transmission rates between 56 kilobytes per second (kb/s) and 1.544 megabytes per second (Mb/s). Frames may vary in length up to a design limit of approximately 1 kilobyte.

The Telco Carrier Cloud 412 is a communication network which receives the frames destined for the DPC 600 sent by the WAN router 408 from the DACs 400. As is known to persons of ordinary skill in the art, carriers provide communication services at local central offices. These central offices contain networking facilities and equipment to interconnect telephone and data communications to other central offices within its own network and within networks of other carriers.

Since carriers share the component links of the interconnection network, data communication must be dynamically assigned to links in the network according to availability. Because of the dynamic nature of the data routing, the interconnection network is referred to as a carrier cloud of communication bandwidth.

All the DAC 400 equipment is on fully redundant on-line UPS power supplies to insure maximum power availability. Further, to minimize the time for trouble detection, trouble analysis and repair, all the DAC 400 equipment incorporates trouble detection and remote reporting/diagnostics as is known to an artisan of ordinary skill in the art.

FIG. 5 is a flow chart 500 describing the polling of the DATs 200 by a DAC 400 and the transmission of the TECBIs from the DATs 200 to the DAC 400. In step 502, the DAC server 402 reads the address of the first DAT 200 in its region for polling. In step 504, a modem in the modem bank 404 dials the first DAT 200. The DAC 400 determines whether the call

to the DAT 200 was successful in step 506. If the call to the first DAT 200 was unsuccessful, the DAC 400 will record the error condition in the session summary report and will report the error to the DPC 600 in step 522.

If the call to the first DAT 200 was successful, the DAC 400 will verify that the DAT 200 is ready to transmit in step 508. If the DAT 200 is not ready to transmit, the DAC 400 will record the error condition in the session summary report and will report the error to the DPC 600 in step 522.

If the DAT 200 is ready to transmit in step 508, the DAT 200 will transmit a TECBI packet header to the DAC 400 in step 510. The DAC 400 will determine whether the transmission of the TECBI packet header was successful in step 512. If the transmission of the TECBI packet header was unsuccessful, the DAC 400 will record the error condition in the session summary report and will report the error to the DPC 600 in step 522.

If the transmission of the TECBI packet header was successful in step 512, the DAT 200 will transmit a TECBI packet to the DAC 400 in step 514. The DAC 400 will determine whether the transmission of the TECBI packet was successful in step 516. If the transmission of the TECBI packet header was unsuccessful, the DAC 400 will record the error condition in the session summary report and will report the error to the DPC 600 in step 522.

If the transmission of the TECBI packet was successful in step 516, the DAC 400, in step 518, will compare the TECBI packet header transmitted in step 510 to the TECBI packet transmitted in step 514. If the TECBI packet header does not match the TECBI packet, the DAC 400 will record the error condition in the session summary report and will report the error to the DPC 600 in step 522.

If the TECBI packet header matched the TECBI packet in step 518, the DAC 400 will set the status of the TECBI packet to indicate that it is ready for transmission to the DPC 600 in step 520. The DAC 400 will also transmit the status to the DAT 200 to indicate successful completion of the polling

and transmission session in step 520. Next, the DAC 400 will determine whether TECBIs have been transmitted from all of the DATs 200 in its region in step 524. If all DATs 200 in the DAC's 400 region have transmitted TECBIs to the DAC 400, the DAC 400 will compile a DAT 200 status report in step 528 before terminating the session.

If one or more DATs 200 in the DAC's 400 region have not transmitted TECBIs to the DAC 400, the DAC 400 will get the address of the next DAT 200 in the region in step 526. Next, control returns to step 504 where the next DAT 200 in the DAC's 400 region will be polled as previously discussed.

In the preferred embodiment, the DAC server 402 initiates the polling and data transmission at optimum toll rate times to decrease the cost of data transmission. In addition to the raid drives and redundant servers, the DAC 400 will also have dual tape backup units which will periodically backup the entire data set. If there is a catastrophic failure of the DAC 400 , the tapes can be retrieved and sent directly to the DPC 600 for processing. As the DAT 200 polling and data transmission progresses, the DAC 400 will periodically update the DPC 600 with its status. If there is a catastrophic failure with the DAC 400, the DPC 600 would know how much polling and backup has been done by the failing DAC 400. Accordingly, the DPC 600 can easily assign another DAC 400 to complete the polling and data transmission for the DATs 200 in the failed DAC's 400 region.

FIG. 6 is a block diagram of the DPC 600 architecture. The DPC 600 accumulates, processes and stores images for later retrieval by DataTreasury™ System retrieval customers who have authorization to access relevant information. DataTreasury™ System retrieval customers include credit card merchants, credit card companies, credit information companies and consumers. As shown in FIG. 6 and FIG. 1, the DPC 600 polls the DACs 400 and receives TECBIs which have accumulated in the DACs 400.

In the preferred embodiment, the DPC server 602 comprises stand-alone Digital Equipment Corporation (DEC) SMP

Alpha 4100 4/566 servers which are connected on a common network running Windows NT. The DEC Alpha servers manage the collection and intermediate storage of images and data which are received from the DACs 400.

In the preferred embodiment, the DPC server 602 also comprises EMC 3700 SYMMETRIX CUBE Disk Storage Systems, which store the images and data collected and managed by the DEC Alpha servers. Like the DAC 400 architecture, the DPC 600 architecture uses a SYMMETRIX Remote Data Facility (SRDF), available from EMC, to enable multiple, physically separate data centers housing EMC Storage Systems to maintain redundant backups of each other across a Wide Area Network (WAN).

Like the DAC 400 architecture, the DPC 600 architecture uses a WEB based paradigm using an enhanced Domain Name Services (DNS), the Microsoft Component Object Model (DCOM), and Windows NT Application Program Interfaces (APIs) to facilitate communication and load balancing among the servers comprising the DPC server 602 as described above in the discussion of the DAC 400 architecture.

The workstation 604 performs operation control and system monitoring and management of the DPC 600 network. In the preferred embodiment, the workstation 604, available from Compaq, is an Intel platform workstation running Microsoft Windows NT 4.x. The workstation 604 should be able to run Microsoft Windows NT 5.x when it becomes available. The workstation 604 executes CA Unicenter TNG software to perform network system monitoring and management. The workstation 604 executes SnoBound Imaging software to display and process TECBIs.

The workstation 604 also performs identification verification by comparing signature data retrieved remotely by the DATs 200 with signature data stored at the DPC 600. In the preferred embodiment, signature verification software, available from Communications Intelligence Corporation of Redwood Shores, California executing on the workstation 604 performs the identification verification. As is known to

26

persons of ordinary skill in the art, the workstation 604 could execute other software to perform identification verification by comparing biometric data including facial scans, fingerprints, retina scans, iris scans and hand geometry. Thus, the DPC 600 could verify the identity of a person who is making a purchase with a credit card by comparing the biometric data captured remotely with the biometric data stored at the DPC 600.

As is known to persons of ordinary skill in the art, the DataTreasury™ System 100 could use workstations with central processing units from other integrated circuit vendors as long as the chosen workstation has the ability to perform standard operations such as fetching instructions, fetching data, executing the fetched instructions with the fetched data and storing results. Similarly, the DataTreasury™ System 100 could use alternate windows operating systems and network monitoring software as long as the selected software can monitor the status of the workstations and links in the network and display the determined status to the operator.

The Remote Data Entry Gateway 614 and the Remote Offsite Data Entry Facilities 616 correct errors which occurred during data capture by the DAT 200. Since the DataTreasury™ System 100 partitions the document as described in the discussion of the sample receipt of FIG. 3b, the operator at the Remote Data Entry Gateway 614 or the Remote Offsite Date Entry Facilities 616 only needs to correct the portion of the document or image snippet which contained the error.

Partitioning improves system performance, decreases system cost and improves system quality. With partitioning, the DPC Server 602 only sends the portion of the document containing the error to the Remote Data Entry Gateway 614 or the Remote Offsite Data Entry Facilities 616. Since the operator at these data entry locations only sees the portion of the document which contained the error, she can quickly recognize and correct the error. Without partitioning, the operator would have to search for the error in the entire document. With this inefficient process, the operator would

need more time and would be more likely to make a mistake by missing the error or making a modification in the wrong location. Accordingly, partitioning improves system performance and quality by increasing the speed and accuracy of the error correction process.

Similarly, partitioning decreases the traffic on the DPC LAN 606 and the Telco Carrier Cloud 412 because the DPC Server 602 only sends the image snippet containing the error to the Remote Offsite Data Entry Facility 616 or the Remote Data Entry Gateway 614. Accordingly, partitioning decreases system cost by reducing the bandwidth requirement on the interconnection networks.

A DPC LAN 606 facilitates communication among the devices which are connected to the LAN 606 including the DPC server 602 and the network workstation 604. In the preferred embodiment, the DPC LAN 606 uses a switched 100BaseT/10BaseT communication hardware layer protocol like the DAC LAN 406 discussed earlier. In the preferred embodiment, the DPC LAN 406 is a high speed OC2 network topology backbone supporting TCP/IP. The CISCO Catalyst 5500 Network Switch supports the DPC LAN 606 connectivity among the devices connected to the LAN 606.

As is known to persons of ordinary skill in the art, alternate LAN architectures could be used to facilitate communication among the devices of the LAN 406. For example, the LAN 406 could use a hub architecture with a round robin allocation algorithm, a time division multiplexing algorithm or a statistical multiplexing algorithm.

A Wide Area Network (WAN) router 612 connects the DPC LAN 606 to the WAN to facilitate communication between the DACs 400 and the DPCs 600. In the preferred embodiment, the WAN router 612 is a CISCO 7507 WAN Router. The WAN router 612 uses frame relay connectivity to connect the DPC LAN 612 to the WAN. As is known to persons of ordinary skill in the art, alternate devices, such as the NORTEL Magellen Passport "50" Telecommunication Switch, could be used to facilitate communication between the DACs 400 and the DPCs 600 as long

as the selected router meets the performance and quality communication requirements of the system

The DPC 600 has a three tier storage architecture to support the massive storage requirement on the DataTreasury™ System 100. In the preferred embodiment, the storage architecture consists of Fiber Channel RAID technology based EMC Symmetrix Enterprise Storage Systems where individual cabinets support over 1 Terabyte of storage. After TECBI images have been processed and have been on-line for 30 days, they will be moved to DVD based jukebox systems. After the TECBI images have been on-line for 90 days, they will be moved to Write Once Read Many (WORM) based jukebox systems 608 for longer term storage of up to 3 years in accordance with customer requirements.

In an alternate embodiment, the DPC 600 is intended to also configure a High Density Read Only Memory (HD-ROM) when it becomes available from NORSAM Technologies, Los Alamos, New Mexico, into optical storage jukebox systems 610, such as that which is available from Hewlett Packard, to replace the DVD components for increased storage capacity. The HD-ROM conforms to CD-ROM form factor metallic WORM disc. The HD-ROM currently has a very large storage capacity of over 320 giga bytes (320 GB) on a single platter and has an anticipated capacity of several terabytes (TB) on a single platter. The DPC 600 uses IBM and Philips technology to read from the HD-ROM and to write to the HD-ROM.

The DPC Alpha servers of the DPC server 602 insert images and data received from the DACs 400 into a single database which is stored on the Digital Storage Works Systems using a data manipulation language as is well known to persons of ordinary skill in the art. In the preferred embodiment, the database is the V8.0 Oracle relational database which was designed to support both data and image storage within a single repository.

As known to persons of ordinary skill in the art, a relational database consists of a collection of tables which have a unique name. See, e.g., Chapter Three of Database

System Concepts by Korth and Silberschatz. A database schema is the logical design of the database. Each table in a relational database has attributes. A row in a table represents a relationship among a set of values for the attributes in the table. Each table has one or more superkeys. A superkey is a set of one or more attributes which uniquely identify a row in the table. A candidate key is a superkey for which no proper subset is also a superkey. A primary key is a candidate key selected by the database designer as the means to identify a row in a table.

As is well known to persons of ordinary skill in the art, the DataTreasury™ System 100 could use other database models available from other vendors including the entity relationship model as long as the selected database meets the storage and access efficiency requirements of the system. See, e.g., Chapter 2 of Database System Concepts by Korth and Silberschatz.

An exemplary DPC 600 basic schema consists of the tables listed below. Since the names of the attributes are descriptive, they adequately define the attributes' contents. The primary keys in each table are identified with two asterisks (**). Numeric attributes which are unique for a particular value of a primary key are denoted with the suffix, "NO". Numeric attributes which are unique within the entire relational database are denoted with the suffix, "NUM".

I. CUSTOMER: This table describes the DataTreasury™ System customer.

A. **CUSTOMER_ID
B. COMPANY_NAME
C. CONTACT
D. CONTACT_TITLE
E. ADDR1
F. ADDR2
G. CITY
H. STATE_CODE

I. ZIP_CODE
J. COUNTRY_CODE
K. VOX_PHONE
L. FAX_PHONE
M. CREATE_DATE

II. CUSTOMER_MAIL_TO: This table describes the mailing address of the DataTreasury™ System customer.
A. **MAIL_TO_NO
B. **CUST_ID
C. CUSTOMER_NAME
D. CONTACT
E. CONTACT_TILE
F. ADDR1
G. ADDR2
H. CITY
I. STATE_CODE
J. ZIP_CODE
K. COUNTRY_CODE
L. VOX_PHONE
M. FAX_PHONE
N. CREATE_DATE
O. COMMENTS

III. CUSTOMER_DAT_SITE: This table describes the DAT location of the DataTreasury™ System customer.
A. **DAT_SITE_NO
B. **CUST_ID
C. CUSTOMER_NAME
D. CONTACT
E. CONTACT_TILE
F. ADDR1
G. ADDR2
H. CITY
I. STATE_CODE
J. ZIP_CODE
K. COUNTRY_CODE

L. VOX_PHONE
M. FAX_PHONE
N. CREATE_DATE
O. COMMENTS

IV. CUSTOMER_SITE_DAT: This table describes the DAT site(s) of the DataTreasury™ System customer.
A. **DAT_TERMINAL_ID
B. **DAT_SITE_NO
C. **CUST_ID
D. INSTALL_DATE
E. LAST_SERVICE_DATE
F. CREATE_DATE
G. COMMENTS

V. DATA_SPEC: This table provides data specifications for document partitioning and extraction.
A. **DATA_SPEC_ID
B. **CUST_ID
C. DESCR
D. RECORD_LAYOUT_RULES
E. CREATE_DATE
F. COMMENTS

VI. DATA_SPEC_FIELD: This table provides field data specifications for document partitioning and extraction.
A. **DATA_SPEC_NO
B. **DATA_SPEC_ID
C. FIELD_NAME
D. DESCR
E. DATA_TYPE
F. VALUE_MAX
G. VALUE_MIN
H. START_POS
I. END_POS
J. FIELD_LENGTH
K. RULES

L. CREATE_DATE

M. COMMENTS

VII. TEMPL_DOC: This table specifies the partitioning of a predefined document.

A. **TEMPL_DOC_NUM

B. DATA_SPEC_ID

C. DESCR

D. RULES

E. CREATE_DATE

F. COMMENTS

VIII. TEMPL_FORM: This table defines the location of forms on a predefined document.

A. **TEMPL_FORM_NO

B. **TEMPL_DOC_NUM

C. SIDES_PER_FORM

D. MASTER_IMAGE_SIDE_A

E. MASTER_IMAGE_SIDE_B

F. DISPLAY_ROTATION_A

G. DISPLAY_ROTATION_B

H. DESCR

I. RULES

J. CREATE_DATE

IX. TEMPL_PANEL: This table specifies the location of panels within the forms of a predefined document.

A. **TEMPL_PANEL_NO

B. **TEMPL_SIDE_NO

C. **TEMPL_FORM_NO

D. **TEMPL_DOC_NUM

E. DISPLAY_ROTATION

F. PANEL_UL_X

G. PANEL_UL_Y

H. PANEL_LR_X

I. PANEL_LR_Y

J. DESCR

K. RULES

L. CREATE_DATE

X. TEMPL_FIELD: This table defines the location of fields within the panels of a form of a predefined document.

A. **TEMPL_FIELD_NO

B. **TEMPL_PANEL_NO

C. **TEMPL_SIDE_NO

D. **TEMPL_FORM_NO

E. **TEMPL_DOC_NUM

F. DISPLAY_ROTATION

G. FLD_UL_X

H. FLD_UL_Y

I. FLD_LR_X

J. FLD_LR_Y

K. DESCR

L. RULES

M. CREATE_DATE

XI. DAT_BATCH: This table defines batches of documents which were processed during a DAT session.

A. **DAT_BATCH_NO

B. **DAT_SESSION_NO

C. **DAT_SESSION_DATE

D. **DAT_TERMINAL_ID

E. DAT_UNIT_CNT

F. CREATE_DATE

XII. DAT_UNIT: This table defines the unit in a batch of documents which were processed in a DAT session.

A. **DAT_UNIT_NUM

B. **DAT_BATCH_NO

C. **DAT_SESSION_NO

D. **DAT_SESSION_DATE

E. **DAT_TERMINAL_ID

F. FORM_CNT

G. DOC_CNT

34

H. CREATE_DATE

XIII. DAT_DOC: This table defines documents in the unit of documents which were processed in a DAT session.

A. **DAT_DOC_NO

B. **DAT_UNIT_NUM

C. DOC_RECORD_DATA

D. CREATE_DATE

The DATA_SPEC, DATA_SPEC_FIELD, TEMPL_DOC, TEMPL_FORM, TEMPL_PANEL and TEMPL_FIELD tables implement the document partitioning algorithm mentioned above in the discussion of the sample receipt of FIG. 3b. The cross product of the DATA_SPEC and DATA_SPEC_FIELD tables partition arbitrary documents while the cross product of the TEMPL_DOC, TEMPL_FORM, TEMPL_PANEL and TEMPL_FIELD tables partition predefined documents of the DataTreasury™ System 100. The TEMPL-FORM defines the location of forms on a predefined document. The TEMPL-PANEL defines the location of panels within the forms of a predefined document. Finally, the TEMPL_FIELD table defines the location of fields within the panels of a form of a predefined document.

The DPC 600 performs data mining and report generation for a wide variety of applications by returning information from the data base. For example, the DPC 600 generates market trend analysis reports and inventory reports for merchants by analyzing the data from receipts captured by the DAT 200. The DPC 600 also can provide important tax information to the taxpayer in the form of a report or to software applications like tax preparation software by retrieving tax information from the database which originally resided on receipts, documents and electronic transactions captured by the DAT 200. Similarly, the DPC 600 can also provide tax information for particular periods of time for a tax audit.

FIG. 7 is a flow chart 700 describing the polling of the DACs 300 by a DPC 600 and the transmission of the TECBIs from

35

the DACs 300 to the DPC 600. In step 702, the DPC 600 reads the address of the first DAC 300 in its region for polling. In step 704, the DPC 600 connects with a DAC 300 for transmission. The DPC 600 determines whether the connection to the DAC 300 was successful in step 706. If the call to the DAC 300 was unsuccessful, the DPC 600 will record the error condition in the session summary report and will report the error to the DPC 600 manager in step 722.

If the connection to the DAC 300 was successful, the DPC 600 will verify that the DAC 300 is ready to transmit in step 708. If the DAC 300 is not ready to transmit, the DPC 600 will record the error condition in the session summary report and will report the error to the DPC 600 manager in step 722.

If the DAC 300 is ready to transmit in step 708, the DAC 300 will transmit a TECBI packet header to the DPC 600 in step 710. The DPC 600 will determine whether the transmission of the TECBI packet header was successful in step 712. If the transmission of the TECBI packet header was unsuccessful, the DPC 600 will record the error condition in the session summary report and will report the error to the DPC 600 manager in step 722.

If the transmission of the TECBI packet header was successful in step 712, the DAC 300 will transmit a TECBI packet to the DPC 600 in step 714. The DPC 600 will determine whether the transmission of the TECBI packet was successful in step 716. If the transmission of the TECBI packet header was unsuccessful, the DPC 600 will record the error condition in the session summary report and will report the error to the DPC 600 manager in step 722.

If the transmission of the TECBI packet was successful in step 716, the DPC 600, in step 718, will compare the TECBI packet header transmitted in step 710 to the TECBI packet transmitted in step 714. If the TECBI packet header does not match the TECBI packet, the DPC 600 will record the error condition in the session summary report and will report the error to the DPC 600 manager in step 722.

If the TECBI packet header matched the TECBI packet in step 718, the DPC 600 will set the status of the TECBI packet to indicate that it was received at the DPC 600 in step 720. The DPC 600 will also transmit the status to the DAC 300 to indicate successful completion of the polling and transmission session in step 720. Next, the DPC 600 will determine whether TECBIs have been transmitted from all of the DACs 300 in its region in step 724. If all DACs 300 in the DPC's 600 region have transmitted TECBIs to the DPC 600, the DPC 600 will compile a DAC 300 status report in step 728 before terminating the session.

If one or more DACs 300 in the DPC's 600 region have not transmitted TECBIs to the DPC 600, the DPC 600 will get the address of the next DAC 300 in the region in step 726. Next, control returns to step 704 where the next DAC 300 in the DPC's 600 region will be polled as previously discussed.

FIG. 8 is a flow chart 800 describing the data processing performed by the DPC 600. In step 802, the DPC 600 fetches the first TECBI packet. Next, the DPC 600 extracts the first TECBI from the TECBI packet in step 804. In step 806, the DPC 600 inserts the TECBI into the database. In step 808, the DPC 600 extracts the tag header which includes the customer identifier, the encryption keys and the template identifier from the TECBI to obtain the ECBI.

In step 810, the DPC 600 decrypts the ECBI image to obtain the CBI. In step 812, the DPC 600 uncompresses the CBI to obtain the BI. In step 814, the DPC 600 fetches and applies the BI template against the BI. Further the DPC 600 divides the BI into image snippets and tags the BI template with data capture rules in step 814 to form the Tagged Bitmap Image Snippets (TBIS). In step 816, the DPC 600 submits the TBISs for data capture operations to form the IS Derived Data Record (ISDATA). The DPC 600 discards the TBISs upon completion of the data capture operations in step 816. In step 818, the DPC 600 updates the TECBI record in the database with the IS Derived Data.

In step 820, the DPC 600 determines whether it has processed the last TECBI in the TECBI packet. If the last TECBI in the TECBI packet has not been processed, the DPC 600 extracts the next TECBI from the TECBI packet in step 822. Next, control returns to step 806 where the next TECBI will be processed as described above.

If the last TECBI in the TECBI packet has been processed, the DPC 600 determines whether the last TECBI packet has been processed in step 824. If the last TECBI packet has not been processed, the DPC 600 fetches the next TECBI packet in step 826. Next, control returns to step 804 where the next TECBI packet will be processed as described above. If the last TECBI packet has been processed in step 824, the DPC 600 terminates data processing.

As is known to persons of ordinary skill in the art, a user can request information from a relational database using a query language. See, e.g., Chapter Three of Database System Concepts by Korth and Silberschatz. For example, a user can retrieve all rows of a database table having a primary key with particular values by specifying the desired primary key's values and the table name on a select operation. Similarly, a user can retrieve all rows from multiple database tables having primary keys with particular values by specifying the desired primary keys' values and the tables with a select operation.

The DataTreasury™ System provides a simplified interface to its retrieval customers to enable data extraction from its relational database as described in FIG. 9. For example, a DataTreasury™ System customer can retrieve the time, date, location and amount of a specified transaction.

The DPC 600 performs data mining and report generation for a wide variety of applications by returning information from the data base. For example, the DPC 600 generates market trend analysis reports and inventory reports for merchants by analyzing the data from receipts captured by the DAT 200. The DPC 600 also can provide important tax information to the taxpayer in the form of a report or to tax

preparation software by retrieving tax information from the database which originally resided on receipts, documents and electronic transactions captured by the DAT 200. Similarly, the DPC 600 can also provide tax information for particular periods of time for a tax audit.

FIG. 9 is a flowchart 900 describing the data retrieval performed by the DPC 600. In step 902, the DPC 600 receives a TECBI retrieval request. In step 904, the DPC 600 obtains the customer identifier. In step 906, the DPC 600 determines whether the customer identifier is valid. If the customer identifier is not valid, control returns to step 904 where the DPC 600 will obtain another customer identifier.

If the customer identifier is valid in step 906, the DPC 600 will obtain the customer security profile in step 908. In step 910, the DPC 600 receives a customer retrieval request. In step 912, the DPC 600 determines whether the customer retrieval request is consistent with the customer security profile. If the customer retrieval request is not consistent with the customer security profile, control returns to step 910 where the DPC 600 will obtain another customer retrieval request. If the customer retrieval request is consistent with the customer security profile, the DPC 600 will transmit the results to the customer as indicated by the customer security profile in step 914.

FIG. 10 is a flow chart describing the use of the DataTreasury™ system to process checks. In step 1004, the DataTreasury™ system captures the check at the payer's remote location in the preferred embodiment before the payer presents the check to the payee. Alternatively, the payer simply presents or mails the check to the payee. The capture of the check at the payer's remote location in step 1004 enables subsequent comparison of the check as written by the payer with the check as received by the payee. In other words, this step enables the detection of check alteration from fraudulent check schemes where a check is intercepted before receipt by the payee and chemically washed to allow the perpetrator to work with a blank check.

39

In step 1006, the DataTreasury™ system captures the check and the payer's biometric data at the payee's remote location. In an alternate embodiment, the DataTreasury™ system sends electronic transaction data representing the check from the payer's remote location to the payer's remote location. In step 1008, the DataTreasury™ system performs verification of the check and biometric data by comparing the remotely captured data with the data stored at a central location. The validation further includes checking the courtesy amount and the payer's signature.

In step 1010, the DataTreasury™ system determines whether the verification was successful. If the verification of step 1010 was not successful, the system transmits an error message to the remote locations in step 1012 and returns to step 1004 for resubmission. If the verification of step 1010 was successful, the system creates an electronic transaction representing the check at a central location in step 1014. The electronic transaction representing the check consists of the payer bank's identification number, routing information, the payer's account number, a payer's check, a payer bank's draft, the amount of the check or draft, the payee bank's identification number, the payee bank's routing information, and the payee's account number. In step 1016, the electronic transaction representing the check is transmitted to the payee bank. In step 1018, the payee bank transmits the electronic transaction representing the check to the payer bank.

In step 1020, the payer bank verifies the electronic transaction representing the check and determines whether to approve a fund transfer. If the payee bank grants approval in step 1020, the payer bank transfers the funds from the payer bank to the payee bank in step 1022. In step 1024, the DataTreasury™ system notifies the payee bank and the remote locations as to the status of the transfer.

While the above invention has been described with reference to certain preferred embodiments, the scope of the present invention is not limited to these embodiments. One

10

skilled in the art may find variations of these preferred embodiments which, nevertheless, fall within the spirit of the present invention, whose scope is defined by the claims set forth below.